Linear Regression: Equations

Ki Hyun Kim

nlp.with.deep.learning@gmail.com



Objective

• 데이터셋(\mathcal{D})이 주어졌을 때, loss를 최소로 하는 파라미터(θ)를 찾자

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N, \ ext{where } x_{1:N} \in \mathbb{R}^{N imes n} ext{ and } y_{1:N} \in \mathbb{R}^{N imes m}.$$

$$egin{aligned} \hat{ heta} &= rgmin_{ heta \in \Theta} \sum_{i=1}^N \left\| y_i - f_{ heta}(x_i)
ight\|_2^2, \ ext{where } heta &= \{W,b\} ext{ and } f_{ heta}(x) = x \cdot W + b. \end{aligned}$$

$$x_{1:N} = \left[egin{array}{cccc} x_{1,1} & \cdots & x_{1,n} \ dots & \ddots & dots \ x_{N,1} & \cdots & x_{N,n} \end{array}
ight]$$

$$y_{1:N} = egin{bmatrix} y_{1,1} & \cdots & y_{1,m} \ dots & \ddots & dots \ y_{N,1} & \cdots & y_{N,m} \end{bmatrix}$$

Loss Minimization using Gradient Descent

• Loss 함수를 파라미터(W,b)로 미분하여, 기울기 값을 활용해보자

$$\hat{y}_{1:N} = x_{1:N} \cdot W + b, \ ext{where } W \in \mathbb{R}^{n imes m} ext{ and } b \in \mathbb{R}^m.$$

$$egin{align} \mathcal{L}(heta) &= \sum_{i=1}^{N} \|y_i - \hat{y}_i\|_2^2 \ &= \sum_{i=1}^{N} \sum_{j=1}^{m} (y_{i,j} - \hat{y}_{i,j})^2 \ &= \sum_{i=1}^{N} \sum_{j=1}^{m} y_{i,j}^2 - 2y_{i,j} \cdot \hat{y}_{i,j} + \hat{y}_{i,j}^2, \end{aligned}$$

$$W = \begin{bmatrix} w_{1,1} & \cdots & w_{1,m} \\ \vdots & \ddots & \vdots \\ w_{n,1} & \cdots & w_{n,m} \end{bmatrix}$$

Loss Minimization using Gradient Descent - Detail

• Loss 함수를 파라미터(θ)로 미분하여, 기울기 값을 활용해보자

$$\theta = \{W, b\}$$
, where $W \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^m$

$$\theta = \{W, b\}, \text{ where } W \in \mathbb{R}^{n \times m}, b \in \mathbb{R}^m \quad W = \begin{bmatrix} w_{1,1} & \cdots & w_{1,m} \\ \vdots & \ddots & \vdots \\ w_{n,1} & \cdots & w_{n,m} \end{bmatrix}$$

$$egin{aligned} \mathcal{L}(heta) &= \sum_{i=1}^{N} \|y_i - \hat{y}_i\|_2^2 & heta \leftarrow heta - \eta
abla_{ heta} \mathcal{L}(heta) & \downarrow & \\ &= \sum_{i=1}^{N} \sum_{j=1}^{m} (y_{i,j} - \hat{y}_{i,j})^2 & W \leftarrow W - \eta
abla_{W} \mathcal{L}(heta) & \\ &= \sum_{i=1}^{N} \sum_{j=1}^{m} y_{i,j}^2 - 2y_{i,j} \cdot \hat{y}_{i,j} + \hat{y}_{i,j}^2, & W_{k,j} \leftarrow W_{k,j} - \eta rac{d\mathcal{L}(heta)}{dW_{k,j}} & \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} x_{i,k} \times W_{k,j} + b_{j}. & b_{j} \leftarrow b_{j} - \eta rac{d\mathcal{L}(heta)}{db_{j}} & \end{aligned}$$